Analysis of HLA Frequencies in Population Cohorts for Design of HLA-Based HIV Vaccines

Frances E. Ward, Shiangtai Tuan, and Barton F. Haynes

Departments of Immunology, Information Technology (OIT), and Medicine, and the Duke Center for AIDS Research, Duke University Medical Center, Durham, North Carolina 27710

As HIV infection and AIDS continue to spread around the world, the need to develop protective HIV vaccines is urgent. However, the most effective approaches to be taken remain uncertain. Recently, evidence for protective immunity to HIV infection has been suggested from the observation of cellular and humoral anti-HIV immune responses in HIV-infected individuals who have not developed AIDS after about 10 years [1]. While the immune mechanisms that lead to protection from developing AIDS are incompletely understood, recent findings suggest that both cytotoxic (CTL) and helper T (Th) cells may play important roles in anti-HIV protective immunity [2,3]. Moreover, HLA genes or genes linked to or associated with the HLA complex have been suggested to predispose to either rapid progressor (AIDS within approximately 3 years of infection) versus non-progressor status (no progression to AIDS after 10 years) [rev in 4].

HLA genes present peptides to cytotoxic and helper T cells in a restricted manner; HLA-A, -B, and -C class I alleles generally present immunogenic peptides to CTL and HLA class II alleles present immunogenic peptides to Th cells [rev in 5]. These data have given rise to the notion that the nature of the HIV peptides presented by host HLA molecules to CTL and Th cells may determine both the quality and the quantity of host anti-HIV cellular responses. Thus, one approach to design of HIV preventive immunogens that would induce salutary anti-HIV immune responses would be to design HLA-based HIV vaccines following analysis of the HLA alleles present in the cohort to be immunized and analysis of the most common HIV variants present in the geographic location of the cohort [6]. This article discusses some of the HLA-related issues relevant to the design of HLA-based HIV vaccine, and suggests an analytic approach that might be taken if suitable HLA and HIV data bases were established.

The human major histocompatibility complex (MHC), of which the HLA class I and class II genes are part, is characterized by the presence of several multigene families, extensive polymorphism at many loci, and significant linkage disequilibrium between alleles at particular loci. The extent of genetic polymorphism is still being determined but currently at least 59, 126, and 36 different alleles are known for HLA - A, -B, and -C loci, respectively [7]. Similarly, 135, 25, and 65 different alleles have been recognized for the respective HLA class II genes, HLA-DRB1, -DQB1, and -DPB1 [8]. In most populations, a few alleles are frequent (gene frequency >10%) but most occur at low frequency (less than 10%) and a number of the latter may be rare (gene frequency <1%). As is the case for other genetic polymorphisms, the frequency of HLA alleles differs among populations. An allele that is common in one population may be rare in another. Some alleles are limited to particular ethnic populations, while others are widely shared among ethnically distinct populations.

For example, the allele HLA-A36 is found only among individuals having African ancestry [9]; on the other hand, serologically defined HLA-A2 occurs rather frequently in most populations studied world-wide to date [10]. It should be noted, however, that 15 different subtypes of A2 have been identified by molecular methods [7], but the distribution of these variants in different populations is unknown. Using DNA sequencing analysis, recent population studies among South American Indian tribes have revealed new allelic variations, possibly the result of relatively recent mutations and positive selection pressures, that are limited, or unique to, a particular tribal group [11,12].

Linkage disequilibrium, a significant association between two alleles at different loci, has been well documented within the MHC [13]. The association of HLA-A1, -B8, and -DR3 in Caucasian populations is one of the strongest and most widely cited examples of this genetic phenomenon [14] and is an HLA haplotype associated with rapid progression to AIDS in HIV-infected subjects [15]. Many other examples of HLA linkage disequilibrium have been observed in a variety of populations; but the genetic mechanism(s) that maintains HLA linkage disequilibrium remains unknown.

Given that extensive polymorphism is known to exist within the HLA system and that additional genetic variation is likely to be revealed as more populations are studied using molecular methods, is it feasible to develop an HLA-based HIV vaccine? Must one develop a unique vaccine for each ethnically homogeneous population? Or will it be possible to design one that is effective in a heterogeneous population or in several distinct populations? The answer to these questions await appropriate field testing; in this review, we explore the feasibility of developing HLA-based HIV vaccines based on currently available data.

In this edition of the Los Alamos database, Korber et al [16] have compiled an extensive HIV CTL and Th epitope database and summarized relevant HLA restriction elements. Together with the distribution of the HLA alleles in target populations, these data provide information essential to the design of HLA-based HIV vaccines. Obviously, there are obstacles to HIV vaccine development in general and to design of HLA-based HIV vaccines in particular [rev in 17]. The obstacles include HIV variability, mutability, and the disparate geographic distribution of HIV variants. The HLA system provides an additional hurdle to design of HLA-based HIV vaccines. Detailed HLA gene frequency distributions, incorporating the incidence of recently defined alleles, do not yet exist for most human populations. Nor have HLA restriction elements for HIV CTL epitopes been defined for recently identified and rare HLA alleles.

Results obtained in eleven International Histocompatibility Testing Workshops, particularly the most recent ones [18–21], provide a valuable, although incomplete, HLA database for preliminary analyses for HLA-based vaccine immunogen design. Although data from several laboratories were pooled for the International Workshops, generally no attempt was made to obtain frequencies representative of the different ethnic groups; that is, individuals were not sampled randomly. However, more appropriate data may be available shortly for populations in the United States. Currently, federally funded projects to better define histocompatibility antigens in African and native Americans are underway, and a large amount of data collected by the National Bone Marrow Registry is being analyzed for distribution of HLA gene frequencies in different populations. Hopefully, more complete information will be obtained for other populations in the near future.

Recognizing these imperfections in available data, we have used the International Histocompatibility Workshop data [19,21], supplemented with published data from selected laboratories [22,23], as the best currently available estimate of the frequencies of HLA alleles that have been shown to serve as restriction elements for HIV CTL epitopes [16]. Table 1 summarizes these frequencies for the four populations (African Americans, North American Indians, USA Caucasians, and Thais) that we have considered in our analysis of HLA-based peptide vaccines. Section II of the Los Alamos HIV epitope database of Korber et al [16] lists the CTL epitopes by HLA restriction element. Using these two sets of data and the Hardy-Weinberg theorem [24], we have estimated the proportion of each of the four populations that would be predicted to present peptides to the immune system if a limited number of HIV epitopes were included in a vaccine designed specifically for that population. We have also made similar calculations for a vaccine designed to be immunogenic for all four populations. These results are presented in Table 2.

Table 1   **Frequencies of HLA Class I Alleles That are Known to Serve as HIV CTL Restriction Elements in Four Populations**

HLA Alleles	Frequencies*			
	African Americans	USA Caucasians	North American Indians	Thais
A2	16.7	28.3	25.5	25.5
A3	8.9	12.2	2.9	1.5
A11	2.3	5.5	1.0	32.5
A24	4.7	9.6	19.6	14.6
A28	10.9	4.5	6.9	0.8
A30	9.5	2.6	2.0	1.1
A31	1.7	2.0	27.5	1.7
A32	1.0	5.1	2.0	0.2
A33	8.1	1.0	1.0	13.6
B7	8.3	10.0	3.9	2.7
B8	3.2	10.0	5.6	0.2
B12 (44)	6.2	10.4	3.9	5.4
B13	0.9	3.0	1.0	9.3
B14	3.0	4.1	2.9	0.4
B17	10.9	4.9	1.0	8.1
B18	3.3	4.9	1.0	2.5
B27	1.6	4.1	2.9	6.0
B35	7.7	8.5	18.6	2.5
B37	0.9	2.2	0.0	1.4
B52	1.1	1.2	2.9	3.1
B53	12.8	0.8	0.0	0.0
B57	4.2	3.9	1.0	5.2
B60	1.3	4.5	2.9	8.3
B62	1.4	5.5	4.9	5.0
Cw3	9.6	12.6	22.4	15
Cw4	21.0	9.8	15.4	6

*Frequencies for HLA-A and HLA-B alleles are taken from HLA 1991 [21], HLA-C for African Americans and USA Caucasians are taken from Histocompatibility Testing 1984 [19], HLA-C for North American Indians from Williams and McAuley, 1992 [22], and HLA-C for Thais from the ProceedingsoftheSecondAsiaandOceaniaHistocompatibilityWorkshopConference [23].

Table 2 Proportion of each of the four populations that would be predicted to present peptides to the immune system

Population	HLA Restriction Elements Chosen	HIV Protein	Epitope Location	Epitope
a) African Americans	A2, A3, A11, B35	nef	73–82	QVPLRPMTYK
	A28, B14	gp41	583–592	VERYLKDQQL
	A30, B8	gp41	844–863	RRIRQGLERALL
	B17, B37	nef	117–128	TQGYFPQWQUYT
	Cw4	gp120	576–383	(S)FNCGGEFF

(Proportion of African Americans expected to present these 5 epitopes is 92.3%)

Population	HLA Restriction Elements Chosen	HIV Protein	Epitope Location	Epitope
b) USA Caucasians	A2, A3, A11, B35	nef	73–82	QVPLRPMTYK
	A30, B8	gp41	844–863	RRIRQGLERALL
	B7	gp120	302–312*	RPNNNTRKSI
		nef	126–138*	NYTPGPGVRYPLT
	B12	p24	169–184	IPMFSALSEGATPQDL

(Proportion of USA Caucasians expected to present these 4 epitopes is 90.2%)

Population	HLA Restriction Elements Chosen	HIV Protein	Epitope Location	Epitope
c) North American Indians	A2, A3, A11, B35	nef	73–82	QVPLRPMTYK
	A24	gp41	584–591*	YLKDQQL
		nef	120–144*	YFPDWQNYTPGPGIRYPLTFGWCYK
	A31	gp41	770–780	RLRDLLLIVTR

(Proportion of North American Indians expected to present these 3 epitopes is 96.4%)

Population	HLA Restriction Elements Chosen	HIV Protein	Epitope Location	Epitope
d) Thais	A2, A3, A11, B35	nef	73–82	QVLRPMTYK
	A24	gp41	584–591*	YLKDQQL
		nef	120–144*	YFPDWQNYTPGPGIRYPLTFCGWCYK

(Proportion of Thais expected to present these 2 epitopes is 93.6%)

Population	HLA Restriction Elements Chosen	HIV Protein	Epitope Location	Epitope
e) African Americans USA Caucasians North American	A2, A3, A11, B35	nef	73–82	QVPLRPMTYK
	A28, B14	gp41	583–592	VERYLKDQQL
Indians	A30, B8	gp41	844–863	RRIRQGLERALL
Thais	B17, B37	nef	117–128	TQGYFPQWQNYT
	Cw4	gp120	376–383	(S)FNCGGEFF
	B7	gp120	302–312*	RPNNNTRKSI
		nef	126–138*	NYTPGPGVRYPLT
	B12	p24	169–184	IPMFSALSEGATPQDL
	A31	gp41	770–780	RLRDLLLIVTR
	A24	gp41	584–591*	YLKDQQL
		nef	120–144*	YFPDWQNYTPGPGIRYPLTFCGWCYK

(Proportions of African Americans, USA Caucasians, North American Indians, and Thais expected to present these 9 epitopes are 95.4%, 97.5%, 99.4%, and 97.2%, respectively)

*The criteria upon which choices among peptides should be made are not yet known. It may be important to choose peptides that have been reported to be immunogenic in non-progressors to AIDS or that have been reported to induce immunodominant anti-HIV T-cell responses.

The strategy we have used in this analysis is to first identify the most frequent restriction elements in the population under consideration for vaccination (or common to the 4 populations), to identify peptides that are presented by more than one HLA allele, and then to seek commonality between these two lists. Probability calculations then utilize the frequencies of the commonality alleles supplemented by those of additional high frequency alleles in the population. Alleles are added until the proportion of the individuals in the population carrying one or more of the alleles in the list is at an acceptable level, greater than 90% in our examples. The aim here is to maximize the sum of the HLA gene frequencies that would recognize the least number of different HIV peptides to be included in an HIV immunogen. The next step is to choose the peptides associated with the restricting allele. In some instances only one peptide is associated with an allele while in others, multiple peptides are presented by the same allele.

The appropriate criteria for choosing which immunogenic epitopes should be included in a preventive HIV immunogen are not yet known, but possible criteria for consideration are listed below.

1. Peptides reported to be immunogenic in situations thought to reflect protection from retroviral infection or protection from retroviral-induced immunodeficiency disease (i.e., in non-progressors to AIDS).

2. Peptides presented to the immune system by HLA restricting elements reported to be associated with non progression to AIDS [rev in 4].

3. Peptides reported to be "immunodominant" stimulators of HLA class I-restricted anti-HIV CTL responses [25].

4. Peptides reported presented by several disparate HLA class I allotypes.

For the four population cohorts considered in this analysis, as few as 2 and as many as 5 epitopes are required to achieve a theoretical protection level of at least 90% (Table 2). The different numbers of required epitopes reflect the relative amounts of HLA Class I polymorphism observed in the different ethnic groups and presentation of a peptide by multiple HLA class I molecules. Another factor is the limited number of HLA restriction elements-HIV CTL epitope pairs that have been described. To date, HIV peptides have been associated only with HLA restriction elements that are infrequent in some populations. As more data are accumulated for other epitopes, some that are associated with higher frequency restriction elements may be identified.

A comparison between the individual and combined populations (Table 2) demonstrates that one gains relatively little by including epitopes that are associated with low frequency alleles. The proportion of individuals protected approaches 100% asymptotically so that even adding on epitopes associated with high frequency alleles adds little to the proportion as this level is approached. This is illustrated by the North American Indians where including 6 more epitopes associated with 5 very low frequency alleles and one intermediate frequency allele in the combined theoretical vaccine added only 3.0% protection.

HLA-based HIV vaccine design is presently hampered by incomplete information regarding distributions of HLA alleles in different populations. Therefore, the vaccine design data presented here must be viewed with certain reservations. HLA allele frequencies used in calculations do not account for the many allelic variants that have been discovered recently; and although it is unknown at present whether or not all related variants can present the same HIV peptide, it seems unlikely. This is not a trivial problem, as illustrated in Table 3 where the number of currently recognized molecular variants is tabulated for those HLA class I alleles now known to present at least one HIV CTL epitope. Some of the more frequent alleles (e.g., HLA-A2, -B35, -B62) are known to include more than a dozen variants. The distribution of allelic variants in different populations is only beginning to be determined [26].

Table 3   Number of Closely Related Sequence Variants Currently Identified Worldwide [7]

HLA Allele	Variants	HLA Allele	Variants	HLA Allele	Variants
A2	15	B7	4	Cw3	3
A3	2	B8	2	Cw4	2
A11	2	B12 (44)	6		
A24	7	B13	3		
A28	4	B14	2		
A30	5	B17	6		
A31	2	B18	2		
A32	1	B27	8		
A33	3	B35	13		
		B37	1		
		B52	2		
		B53	1		
		B57	3		
		B60	10		
		B62	28		

Estimates of the proportion of the population able to respond to an HLA-based vaccine may be further complicated by linkage disequilibrium between certain HLA-A, -B, and -C locus alleles and lack of Hardy-Weinberg equilibrium. Use of HIV epitopes restricted by HLA alleles that are in strong linkage disequilibrium would reduce the proportion of individuals responding to the vaccine. In spite of deficiencies in our knowledge of HLA immunogenetics and of HIV epitope presentation to the immune system, a limited number of peptides appear to be sufficient for presentation to the majority of a population cohort.

These analyses represent the initial considerations for design of HIV HLA-based vaccines that might induce salutary anti-HIV Th and CTL responses. It is important to point out that the correlates of protective immunity to HIV are not yet known, and specifically, the need for anti-HIV neutralizing antibodies for protective anti-HIV immune responses is controversial [rev in 4]. Nevertheless, the database of Korber et al [16] provides a new and powerful tool with which to begin to address these complex issues.

References

[1] Sheppard HW, Lang W, and Ascher MS. The characterization of non-progressors: long-term HIV-1 infection with stable CD4+ T-cell levels. AIDS 7:1159–66, 1993.

[2] Pantaleo G, Menzo S, Vaccarezza M, et al. Studies in subjects with long-term nonprogressive human immunodeficiency virus infection. *N. Eng. J. Med.* **332**:209–16, 1995.

[3] Cao Y, Qin L, Zhang L, and Ho DD. Virologic and immunologic characterization of long-term survivors of human immuno-deficiency virus type 1 infection. *N. Eng. J. Med.* **332**:201–8, 1995.

[4] Haynes BF, Pantaleo G, and Fauci AS. Toward an understanding of the correlates of protective immunity to HIV infection. *Science*, In press.

[5] Germain RN and Marguilies DH. The biochemistry and cell biology of antigen processing and presentation. *Ann. Rev. Immunology* **11**:403–50, 1993.

[6] Haynes BF. An appraisal of the current status of HIV vaccine development. *Lancet*, In press.

[7] Arnett KL and Parham P. HLA class I nucleotide sequences, 1995. *Tissue Antigens* **46**:217–57, 1995.

[8] Marsh SGE and Bodmer JG. HLA class II region nucleotide sequences, 1995. *Tissue Antigens* **46**:258–80, 1995.

[9] Chandanayingyong D, Cheung KWK, Davit CS, et al. Report of antigen society no. 101: HLA-A1, -A3, -A36, -A11, -A new. In HLA 1991 Proceedings of the Eleventh International Histocompatibility Workshop and Conference. Tsuhi K, Aizawa M, and Sasazuki T, eds. V1, p 280–85, 1992.

[10] Marsh SGE, Fauchet R, Yang EK, et al. Report of antigen society no. 102: HLA-A2, -A68, -A69, -A9, -A23, -A24. In HLA 1991 Proceedings of the Eleventh International Histocompatibility Workshop and Conference, Tsuji K, Aizawa M, and Sasazuki T, eds. V1, p 285–89, 1992.

[11] Belich MP, Madrigal JA, Hildebrand WH, et al. Unusual HLA-B alleles in two tribes of Brazilian Indians. *Nature* **357**:326–29, 1992.

[12] Watkins DI, McAdams SN, Liu X, et al. New recombinant HLA-B alleles in a tribe of South American Amerindians indicate rapid evolution of MHC class I loci. *Nature* **357**:329–33, 1992.

[13] Ward FE and Amos DB. Human leukocyte antigens and disease. In *Human Immunogenetics— Basic Principles and Clinical Relevance*, Litwin SD, ed. Marcel Dekker, Inc., New York, p 359–78, 1989.

[14] Baur MP and Danilous JA. Population analysis of HLA-A, B, C, DR and other genetic markers. In *Histocompatibility Testing 1980*. Terasaki PI, ed. UCLA Tissue Typing Laboratory, Los Angeles, p 955–93, 1980.

[15] Steele CM, Ludlam CA, Beatson D, et al. HLA haplotype A1, B8, DR3 as a risk factor for HIV-related disease. *Lancet* i, 1185–88, 1988.

[16] HIV Molecular Immunology Database, 1995. Korber B, Koup R, Walker B, Haynes B, Moore J, Myers G (eds). Los Alamos National Laboratory: Published by Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM 87545.

[17] Haynes BF. Scientific and social issues of human immunodeficiency virus vaccine development. *Science* **260**:1279–86, 1993.

[18] *Histocompatibility Testing 1980*. Teresaki PI, ed., UCLA Tissue Typing Laboratory, Los Angeles, 1980.

[19] *Histocompatibility Testing 1984*. Albert ED, Baur MP, and Mayr WR, eds., Springer-Verlag, Berlin, 1984.

[20] *Immunobiology of HLA*, 2 volumes, Dupont B, ed. Springer-Verlag, New York, 1989.

[21] *HLA 1991*, 2 volumes, Tsuji K, Aizawa M, and Sasazuki T, eds., Oxford University Press, Oxford, 1992.

[22] Williams RC and McAuley JE. HLA class I variation controlled for genetic admixture in the Gila River Indian Community of Arizona: a model for the Paleo-Indians. *Human Immunol*. **33**:39–46, 1992.

[23] Chandanayingyong D and Chiewsilp P. HLA antigens in Thais. In Proceedings of the Second Asia and Oceania Histocompatiblity Workshop Conference, Simons MJ and Tait BD, eds., Immunopublishing, Toorak, p 276–87, 1983.

[24] Hardy GH. Mendelian proportions in a mixed population. *Science* **28**:49–50, 1908.

[25] Nowak MA, May RM, Phillips RE, Rowland-Jones S, and Lalloo DG. Antigenic oscillations and shifting immunodominance in HIV-1 infections. *Nature* **375**:606–11, 1995.

[26] Ishikawa Y, Tokunaga K, Kashiwask K, Akaza T, Tadokora K, and Juji T. Sequence-based typing of HLA-A2 alleles using a primer with an extra base mismatch. *Human Immunol*. **42**:315–18, 1995.